

Analisis Sentimen Program Makan Bergizi Gratis Siswa SMAN 01 Manokwari dengan *Naïve Bayes*

Sentiment Analysis of the Free Nutritional Meal Program for Students of SMAN 01 Manokwari Using Naïve Bayes

Ridho Nur Fauzi¹, Thoriq N.F Wajo², Shafril Dwi Firmansyah³, Reason Dominik Harewan⁴,
Iqbal Dwi Wijanarko⁵.

^{1,2,3,4,5} Teknik Informatika, Fakultas Teknik, Universitas Papua

Jl. Gunung Salju Amban, Manokwari, Papua Barat, Indonesia

Email: ridho.nurfauzi123@gmail.com¹, thorixrover@gmail.com², shafrilkeren@gmail.com³, dominik.hrwn@gmail.com⁴,
iqbald.w.899@gmail.com⁵

Abstract

The Free Nutritious Meal Program (MBG) is a government policy aimed at improving students' nutritional intake and supporting the learning process. However, the program's success should not only be assessed from a medical perspective but also from the perception of students as direct recipients. This study aims to analyze the sentiment of students at SMAN 01 Manokwari toward the effectiveness of the MBG program using the *Naïve Bayes* classification method. A total of 252 student opinions were collected and manually labeled into three sentiment categories: positive, neutral, and negative. The data distribution showed 74.2% positive, 15.1% neutral, and 10.7% negative sentiment. Feature extraction was performed using the *TF-IDF* technique, and the dataset was split into 80% training data and 20% testing data. The classification results showed an accuracy of 86.27%, with the highest precision of 0.92, recall of 0.97, and *F1-score* of 0.95 in the positive class. The *AUC* score reached 0.98, indicating excellent classification performance. These results demonstrate that most students responded positively to the MBG program, and the *Naïve Bayes* algorithm is effective in sentiment analysis of student opinions, making it a valuable tool for evaluating future education policies.

Keywords: Sentiment Analysis, *Naïve Bayes*, *TF-IDF*, Free Nutritious Meal Program.

Abstrak

Program Makan Bergizi Gratis (MBG) merupakan kebijakan pemerintah yang bertujuan meningkatkan gizi siswa dan mendukung proses belajar. Namun, keberhasilan program tidak hanya diukur dari aspek medis, tetapi juga dari persepsi siswa sebagai penerima langsung. Penelitian ini bertujuan untuk menganalisis sentimen siswa SMAN 01 Manokwari terhadap efektivitas program MBG menggunakan metode klasifikasi *Naïve Bayes*. Sebanyak 252 data opini siswa dikumpulkan dan dilabeli ke dalam tiga kategori sentimen: positif, netral, dan negatif. Distribusi data menunjukkan 74,2% sentimen positif, 15,1% netral, dan 10,7% negatif. Data diproses menggunakan teknik *TF-IDF* untuk ekstraksi fitur, lalu dibagi menjadi 80% data latih dan 20% data uji. Hasil klasifikasi menunjukkan akurasi 86,27% dengan *precision* tertinggi 0,92, *recall* 0,97, dan *F1-Score* 0,95 pada kelas positif. Nilai *AUC* sebesar 0,98 menunjukkan performa model yang sangat baik. Hasil ini menunjukkan mayoritas siswa merespon positif program MBG dan algoritma *Naïve Bayes* cukup efektif dalam analisis sentimen opini siswa, sehingga dapat digunakan sebagai bahan evaluasi kebijakan pendidikan di masa depan.

Kata kunci: analisis sentimen, *naïve bayes*, *TF-IDF*, program makan bergizi gratis.

1. Pendahuluan

Program makan bergizi gratis merupakan kebijakan pemerintah yang bertujuan meningkatkan kualitas gizi siswa, karena status gizi yang baik berpengaruh pada perkembangan kognitif, terutama bagi anak usia sekolah. Masalah seperti *stunting* akibat kekurangan gizi kronis terbukti berdampak negatif pada fungsi akademik

dan kognitif anak, serta menurunkan produktivitas di masa dewasa [2].

Program Makan Bergizi Gratis dimulai dilaksanakan pada Januari 2025 dan menjadi perhatian publik, terutama di wilayah dengan ketimpangan sosial tinggi, seperti Kabupaten Manokwari, Papua Barat. SMA Negeri 01 Manokwari dipilih sebagai sekolah dengan jumlah siswa besar dan latar belakang sosial ekonomi yang beragam, sehingga menjadi representatif untuk mengkaji dampak sosial dari program Makan Bergizi

Gratis. Program ini diyakini pemerintah dapat meningkatkan status gizi siswa, memperbaiki Indeks Massa Tubuh (IMT), kadar *hemoglobin*, dan menurunkan prevalensi anemia [3].

Namun keberhasilan penerapan program ini tidak bisa hanya dilihat dari sisi fisiologis dan medis, tetapi juga penting untuk mengetahui respon dari sisi siswa yang terdampak dari program ini terhadap pelaksanaan program Makan Bergizi Gratis. Apakah respon para siswa positif, netral, atau negatif. Untuk mengetahui respon tersebut penelitian ini menggunakan pendekatan analisis sentimen, yaitu sebuah teknik untuk mengidentifikasi dan mengklasifikasikan sebuah pendapat dalam bentuk teks, berguna dalam memahami persepsi dari masyarakat terhadap suatu fenomena yang ada, tetapi dalam konteks ini analisis sentimen diimplementasikan kepada respon siswa terhadap program Makan Bergizi Gratis yang diadakan di sekolah mereka. Penelitian ini menggunakan metode *Naïve Bayes*. karena algoritma ini terbukti efektif dalam mengklasifikasikan sebuah teks dengan akurasi yang baik dengan komputasi yang cepat, algoritma *Naïve Bayes* cocok juga digunakan dalam data berukuran kecil hingga menengah seperti data yang didapatkan dari kuesioner atau tanggapan bebas, serta algoritma ini sederhana tapi mampu menangani ketidakseimbangan data sentimen dengan cukup baik.

Menggunakan pendekatan kualitatif dan algoritmik analisis, penelitian ini memiliki tujuan dapat memberikan evaluasi awal terhadap program Makan Bergizi Gratis di sekolah-sekolah, sekaligus memberikan sebuah masukan untuk pemerintah dalam perbaikan kebijakan serupa nantinya, hasil dari penelitian ini diharapkan dapat menjadi bahan referensi bagi penelitian kebijakan publik di bidang sosial dan pendidikan, khusus di daerah seperti Manokwari.

2. Tinjauan Pustaka

2.1 Program Makan Bergizi Gratis

Program Makan Bergizi Gratis (MBG) dimulai sejak tanggal 6 Januari 2025 dan diselenggarakan di 26 provinsi Indonesia dengan menargetkan siswa-siswi dari PAUD hingga SMA dan juga ibu hamil dan menyusui. Program MBG merupakan sebuah program yang dicetus pada masa pemerintahan Prabowo Subianto, yang memiliki beberapa tujuan untuk mewujudkan visi Indonesia Emas 2045 dengan cara membangun sumber daya unggulan, menurunkan angka *stunting*, menurunkan angka kemiskinan, dan menggerakkan ekonomi masyarakat. Dengan begitu dapat menciptakan generasi emas dari bonus demografi yang membuat Indonesia dapat menjadi negara maju serta dapat memberikan manfaat bagi warga yang menerima program ini.

2.2 Analisis Sentimen

Analisis sentimen atau disebut juga *opinion mining* adalah proses menganalisis sebuah teks digital untuk menentukan apa nada emosional dari sebuah pesan teks apakah positif, netral, atau negatif serta dengan mencari pola kata-kata yang digunakan, Tujuan dari analisis sentimen adalah untuk memahami opini, emosi, atau sikap seseorang yang diutarakan dalam bentuk teks. Dalam konteks pendidikan, analisis sentimen telah diterapkan untuk mengevaluasi persepsi siswa terhadap berbagai aspek pembelajaran. [6] melakukan analisis sentimen terhadap kuliah daring menggunakan algoritma *Naïve Bayes*.

2.3 Naïve Bayes

Naïve Bayes adalah algoritma berbasis probabilitas sederhana dan dirancang agar dapat digunakan dengan sebuah asumsi antar variabel penjelas saling bebas atau independen, pembelajaran lebih ditekankan pada pengestimasi probabilitas. Menurut [9], *Naïve Bayes* efektif dalam menangani dataset berukuran besar dan bekerja baik pada analisis sentimen karena mengasumsikan independensi antar fitur. Keuntungan menggunakan *Naïve Bayes* yaitu akurasi dan kecepatannya lebih tinggi pada saat diaplikasikan ke dalam dataset yang jumlahnya banyak.

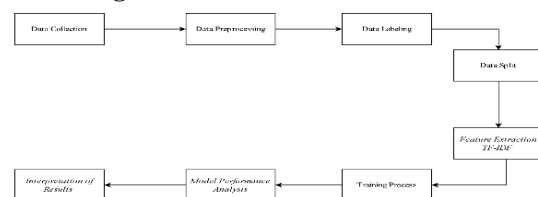
2.4 Penelitian Terkait

Beberapa penelitian sebelumnya telah menunjukkan efektivitas algoritma klasifikasi dalam menganalisis opini publik:

- Anggraini et al. (2022) menggunakan *K-Nearest Neighbor* dan *Decision Tree* untuk mengklasifikasikan respon mahasiswa terhadap pembelajaran daring dan mencapai akurasi lebih dari 80%.
- Hartati & Kusumawardani (2021) menerapkan *Naïve Bayes* untuk menganalisis sentimen masyarakat terhadap vaksinasi COVID-19 dari media sosial.
- Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode *Naïve Bayes*. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*.

Penelitian ini mengadopsi pendekatan serupa namun fokus pada evaluasi kebijakan sosial di bidang pendidikan, yakni program makan bergizi gratis, khususnya di lingkungan siswa SMA dengan latar belakang sosial ekonomi yang beragam.

3. Metodologi Penelitian



Gambar 1 Tahapan Penelitian

Pengambilan data akan menggunakan kuesioner dengan jenis data yaitu kualitatif yang nantinya akan diproses dengan pendekatan data mining yang berfokus pada pengklasifikasian respon dari setiap siswa terhadap program makan bergizi gratis di SMA Negeri 01 Manokwari dengan label positif, netral, dan negatif. Berikut merupakan tahapan yang dilakukan dalam proses penelitian ini:

3.1 Data Collection

Proses pengumpulan data untuk dijadikan sebagai data primer yang akan dilakukan melalui penyebaran kuesioner kepada para siswa SMA Negeri 01 Manokwari sebagai respon untuk program makan bergizi gratis pada penelitian ini. Penyebaran kuesioner akan menggunakan *Google Form* untuk dapat menjangkau siswa secara efisien. Data yang terkumpul nantinya akan dikompilasi dan disimpan dalam format *spreadsheet* agar dapat diproses lebih lanjut menggunakan bahasa pemrograman *python*. Manfaat dari menggunakan *data collection* sendiri salah satunya adalah dari segi kevalidan pada penggunaan data, sehingga data bisa dijadikan sebagai bahan untuk penelitian ilmiah diantaranya adalah meningkatkan kepuasan responden dan mengetahui seberapa besar kepuasan responden melalui *survey*.

3.2 Data Preprocessing

Merupakan proses pengolahan data untuk memastikan bahwa data yang akan dianalisis sudah bersih, terstruktur, dan juga siap digunakan dalam model klasifikasi. Proses ini penting agar hasil lebih akurat dan dapat dipercaya. Penelitian ini akan melalui beberapa tahapan. *Case Folding*, mengubah semua huruf dalam teks menjadi huruf kecil dengan tujuan untuk menyamakan format kata agar tidak dianggap berbeda oleh sistem hanya karena perbedaan huruf besar dan kecil sehingga hasil dari analisis lebih konsisten. *Cleaning*, merupakan proses untuk membersihkan nilai kosong (*missing values*), duplikasi, atau isian tidak valid. Nantinya jika ada data yang tidak lengkap akan dihapus atau diperbaiki agar tidak mempengaruhi hasil klasifikasi dan juga menggunakan *regular expression* untuk pembersihan teks agar nantinya lebih informatif dan model dapat belajar pola dengan baik.

Tokenizing, memecah teks menjadi bagian-bagian kecil yang disebut token, biasanya berupa kata atau frasa dengan tujuan agar tes lebih mudah dianalisis oleh algoritma setiap kalimat nantinya akan diubah menjadi kumpulan kata sehingga dapat diklasifikasikan secara efisien. *Stopword Removal*, menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis dengan tujuan untuk menyederhanakan teks dan hanya mempertahankan kata-kata yang memiliki informasi yang relevan. *Stemming* proses mengubah kata berimbuhan menjadi bentuk dasarnya dengan begitu nantinya dapat menyamakan setiap kata yang

memiliki makna inti yang sama agar lebih mudah dianalisis oleh algoritma [17].

3.3 Data labeling

Merupakan proses pemberian label teks, seperti positif, netral, dan negatif pada data hasil dari *preprocessing*. Misalnya label mungkin menunjukkan apakah data positif, netral maupun negatif, kata-kata apa yang diucapkan dalam kuesioner. Pelabelan data diperlukan untuk berbagai kasus penggunaan termasuk penglihatan komputer, pemrosesan, bahasa alami, dan pengenalan ucapan. Cara kerja metode ini lebih praktis dikarenakan menggunakan pembelajaran yang diawasi melalui algoritma *supervised* untuk memetakan satu *input* ke satu *output*. Agar pembelajaran yang diawasi berfungsi, maka memerlukan *set data* berlabel yang dapat dipelajari model untuk membuat keputusan yang benar.

3.4 Split Data

Merupakan proses pembagian pada dataset ke dalam dua bagian utama, diantaranya yaitu *data training* dan *data testing*, dengan rasio 80:20. Sebanyak 80% data digunakan untuk melatih model (*training*) yang dimana *data training* digunakan untuk melatih dan mengembangkan model. Kumpulan *data training* biasanya digunakan untuk mengestimasi parameter yang berbeda untuk membandingkan kinerja model yang berbeda. Sedangkan 20% sisanya digunakan untuk menguji performa model (*testing*) yang dimana data ini digunakan setelah proses training selesai. Tujuan dari pembagian ini diantaranya adalah untuk memastikan bahwa model tersebut dapat mempelajari pola dari data yang telah tersedia dan diuji kemampuannya dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya secara akurat dan objektif [12].

3.5 Feature Extraction TF-IDF

Feature Extraction TF-IDF adalah teknik pengambilan ciri teks atau feature yang akan dilakukan ekstraksi ke dalam bentuk numerik dan dilakukan analisis [16]. Angka yang berbentuk vektor akan digunakan untuk menentukan bobot dari sebuah kata ke dalam sebuah dokumen atau korpus. Bobot ini berguna untuk menentukan seberapa penting kata tersebut dalam sebuah dokumen. Pada dasarnya untuk perhitungan atau rumus *TF-IDF* terbagi menjadi dua yaitu (*Term Frequency*) menghitung frekuensi jumlah kemunculan kata pada sebuah dokumen dan *IDF* (*Inverse Document Frequency*) nilai untuk mengukur seberapa penting sebuah kata, dengan beracuan semakin kecil nilai *IDF* maka akan dianggap semakin tidak penting kata tersebut.

3.6 Training Process

Merupakan tahap pelatihan model dengan menggunakan data training yang telah dipisahkan sebelumnya. Pada

tahap ini, model algoritma klasifikasi seperti *Naïve Bayes* dilatih untuk mengenali pola dari data kualitatif yang telah diberi label agar dapat memprediksi hasil klasifikasi pada data baru secara akurat dengan cara mengasumsikan bahwa fitur tidak memiliki keterikatan dengan fitur lain walaupun di kelas yang sama[14].

3.7 Model Performance Analysis

Setelah model dilatih, dilakukan evaluasi terhadap kinerja model menggunakan data testing. Beberapa matrik evaluasi yang digunakan antara lain menggunakan *confusion Matrix*, *Area Under Curve*, akurasi, *Precision*, *Recall*, dan *F1-Score* yang disatukan dalam tabel *report* untuk nantinya mempermudah dalam mengevaluasi hasil dari klasifikasi terhadap data yang belum pernah dilihat oleh model sebelumnya. [16].

3.8 Interpretation of Results

Dilakukan interpretasi hasil agar dapat dijadikan sebagai dasar pengambilan keputusan atau rekomendasi terhadap kebijakan yang dikaji, dalam hal ini program makan bergizi gratis di SMA Negeri 01 Manokwari [15].

4. Hasil dan Pembahasan

Pada bagian keempat merupakan tahap pendokumentasian dari setiap langkah dalam proses penelitian yang telah dilakukan. Penelitian ini menggunakan bahasa pemrograman *Python* dengan sejumlah *library* seperti *Scikit-learn*, *Matplotlib*, *NLTK*, *Pandas*, *Regular Expression*, Sastrawi dan *NumPy*. *Library-library* tersebut dimanfaatkan untuk menjalankan seluruh tahapan penelitian, mulai dari data *preprocessing* hingga ke tahap klasifikasi.

4.1 Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan secara *online* dengan memanfaatkan *Google Form* sebagai media utama untuk menyebarkan kuesioner. Proses distribusi dilakukan melalui bantuan wali kelas masing-masing, yang kemudian membagikan tautan formulir kepada para siswa dan siswi SMA Negeri 1 Manokwari. Pemilihan metode *online* didasarkan pada kemudahan akses serta efektivitasnya dalam menjangkau seluruh responden dengan cara yang praktis dan efisien. Melalui pendekatan ini, siswa dapat mengisi kuesioner dengan fleksibel sesuai waktu luang mereka, tanpa harus megananggu oleh jadwal sekolah atau kehadiran fisik. Di sisi lain, peneliti juga memperoleh kemudahan dalam mengumpulkan dan merekap data secara sistematis.

4.2 Data Preprocessing

Tahap *preprocessing* merupakan tahap awal yang bertujuan untuk membersihkan dan merapikan dataset

sebelum masuk ke proses analisis lebih lanjut. Proses ini dilakukan dengan menggunakan bahasa pemrograman *Python* serta didukung oleh sejumlah *library* yang relevan di setiap tahapan. Pada tahap *cleaning*, dilakukan penghapusan karakter yang tidak diperlukan seperti simbol, angka, dan tanda baca menggunakan *library Regular Expression*. Selanjutnya, proses *tokenizing* dilakukan dengan bantuan *library NLTK (Natural Language Toolkit)*, yang memecah kalimat menjadi kata-kata terpisah agar nantinya mempermudah model untuk proses serta mempelajari pola dari data, lalu selanjutnya *stopwords* menyaring kata-kata yang tidak relevan atau umum menggunakan sastrawi. Tahap *stemming*, yaitu mengubah kata ke bentuk dasarnya, juga menggunakan fungsi dari *library* sastrawi. Seluruh tahapan ini dilakukan secara sistematis untuk memastikan bahwa data yang akan digunakan dalam proses klasifikasi berada dalam kondisi bersih dan siap olah. Perubahan data pada setiap tahapan dapat dilihat pada Tabel 1, yang menunjukkan transformasi dataset sesuai dengan fungsi-fungsi yang diterapkan dalam *Python*.

4.3 Data Labeling

Setelah seluruh data melalui proses pembersihan dan standarisasi, langkah berikutnya dalam penelitian ini adalah pemberian label pada masing-masing respon siswa. Tujuan dari pelabelan ini adalah untuk mengelompokkan opini siswa ke dalam kategori tertentu, yang nantinya digunakan dalam proses klasifikasi dengan algoritma *Naïve Bayes*. Pelabelan dilakukan secara manual dengan mempertimbangkan konteks setiap jawaban siswa. Penelitian ini menggunakan tiga kategori label, positif untuk pernyataan siswa yang secara jelas menunjukkan dukungan terhadap program, netral untuk pernyataan yang tidak menunjukkan dukungan maupun penolakan secara jelas terhadap program, atau hanya berisi informasi tanpa opini, dan negatif untuk pernyataan tanggapan yang menunjukkan ketidakpuasan. Setelah seluruh data melalui proses pembersihan dan standarisasi, langkah berikutnya dalam penelitian ini adalah pemberian label pada masing-masing respon

siswa. Tujuan dari pelabelan ini adalah untuk mengelompokkan opini siswa ke dalam kategori tertentu, yang nantinya digunakan dalam proses klasifikasi dengan algoritma *Naïve Bayes*. Pelabelan dilakukan secara manual dengan mempertimbangkan konteks setiap jawaban siswa. Penelitian ini menggunakan tiga kategori *label*, positif untuk pernyataan siswa yang secara jelas menunjukkan dukungan terhadap program, netral untuk pernyataan yang tidak menunjukkan dukungan maupun penolakan secara jelas terhadap program, atau hanya berisi informasi tanpa opini, dan negatif untuk pernyataan tanggapan yang menunjukkan ketidakpuasan.

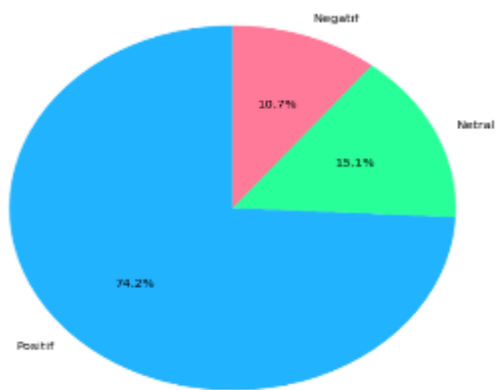
Tabel 1 Data Preprocessing

NO	Nama	Kelas	Respon	Case Folding	Cleaning	Tokenizing	Stemming
1	Syifah An-Nazwa	XII	Saya senang dengan adanya program MBG karena sangat membantu siswa-siswi yang kesulitan membeli makanan. Rasanya enak dan program ini sangat bermanfaat. Untuk perbaikan, mungkin bisa ditambahkan variasi makanan lain agar lebih beragam.	saya senang dengan adanya program mbg karena sangat membantu siswa siswi yang kesulitan membeli makanan rasanya enak dan program ini sangat bermanfaat untuk perbaikan mungkin bisa ditambahkan variasi makanan lain agar lebih beragam	saya senang dengan adanya program mbg karena sangat membantu siswa siswi yang kesulitan membeli makanan dan program ini sangat bermanfaat untuk perbaikan mungkin bisa ditambahkan variasi makanan lain agar lebih beragam	[senang, program, mbg, membantu, siswa, siswi, kesulitan, membeli, makanan, enak, program, bermanfaat, variasi, makanan, beragam]	[senang, program, mbg, bantu, siswa, siswi, sulit, beli, makan, enak, program, manfaat, baik, variasi, makan, ragam]
252	Filgad Sun Life Sitangga ng	XI	Distribusi makanan terkadang memerlukan waktu yang cukup lama, yang dapat mempengaruhi efektivitas jam pelajaran. Selain itu, penting untuk mempertimbangkan dampak alokasi anggaran agar tidak mengurangi dukungan terhadap program lain yang sedang berjalan, seperti kegiatan lomba	distribusi makanan terkadang memerlukan waktu yang cukup lama yang dapat mempengaruhi i efektivitas jam pelajaran selain itu penting untuk mempertimbangkan dampak alokasi anggaran agar tidak mengurangi dukungan terhadap program lain yang sedang berjalan seperti kegiatan lomba	distribusi makanan terkadang memerlukan waktu yang cukup lama yang dapat mempengaruhi efektivitas jam pelajaran selain itu penting untuk mempertimbangkan dampak alokasi anggaran agar tidak mengurangi dukungan terhadap program lain yang sedang berjalan seperti kegiatan lomba	[distribusi, makanan, terkadang, mempengaruhi, efektivitas, jam, pelajaran, mempertimbang kan, dampak, alokasi, anggaran, mengurangi, dukungan, program, berjalan, kegiatan, lomba]	[distribusi, makan, terkadang, pengaruh, efektivitas, jam, ajar, timbang, dampak, alokasi, anggar, kurang, dukung, program, jalan, giat, lomba]

Tabel 2 Hasil Preprocessing

No	Text Preprocessing	Label
1	Syifah An-Nazwa senang program mbg bantu siswa siswi sulit beli makan anak program manfaat baik variasi makan ragam	Positif
2	Marvel lumombo makan anak gizi	Positif
...
...
259	Jekson Romainum makan anak paksa makan hambar kadang mual habis	Negatif
250	Ulvia Syafa Nima Bagu erkadang cocok menu ikan amis wadah makan bau jenis makan terima roti susu ayam menu milik variasi beda	Netral

Langkah berikutnya yaitu melihat jumlah data sentimen positif, netral, dan negatif serta persentasenya dengan bantuan *library Matplotlib*.



Gambar 2 Distribusi Label

Gambar di atas ini menunjukkan visualisasi pada data *labeling* yang bisa kita lihat bahwa terdapat 187 data sentimen positif, 38 data sentimen netral dan 27 data sentimen negatif terhadap Program MBG ini, dengan persentase tersebut yang diperoleh yaitu 74.2% sentimen positif, 15.1% sentimen netral, dan 10.7% sentimen negatif.

4.4 Split Data

Tahapan yang dilakukan sebelum pembobotan *TF-IDF* ini yaitu dengan membagi *dataset* ke dalam data *training* dan data *testing*. Persentase yang digunakan sekitar 80% sebagai *data training* dan 20% sebagai data *testing*. Setelah melalui tahapan *split* data maka dari 252 sentimen, maka dihasilkan bahwa 201 data latih 51 data uji.

4.5 Feature Extraction TF-IDF

Pada tahap *Feature Extraction*, dilakukan pembobotan kata menggunakan *TF-IDF* dengan bantuan fungsi *TfidfVectorizer()* dari *scikit-learn*. *TF-IDF* ini menghitung seberapa sering sebuah kata muncul dalam

dokumen dan seberapa penting kata tersebut dibandingkan dengan dokumen lain.

Tabel 3 Pembobotan TF-IDF

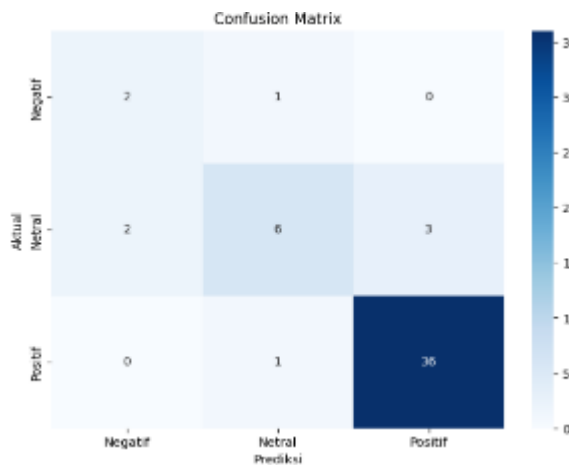
Term	TF-IDF
(0, 608)	0.20359117155989379
(0, 462)	0.281941878603269
(0, 37)	0.17585348213827082
(0, 499)	0.1260350071468872
(0, 246)	0.23165490544227024
(0, 132)	0.16530810204100263
(0, 0)	0.159939893398624
(0, 427)	0.1017924757570852
(0, 322)	0.1803050387142308
(200,585)	0.12542604777989025
(200, 184)	0.15463870420485137
(200, 381)	0.19034803402582934
(200, 330)	0.21956069045079046
(200, 426)	0.19034803402582934
(200, 72)	0.19803196775139575
.....

4.6 Classification Method

Algoritma yang akan digunakan dalam penelitian ini adalah metode *Naïve Bayes* dengan jenis *Naïve Bayes* adalah *Multinomial*. Jenis *Multinomial Naïve Bayes* dipilih karena model ini cocok digunakan untuk data yang berbentuk frekuensi atau jumlah kemunculan kata dalam teks, jenis ini juga bekerja dengan cara mengasumsikan bahwa fitur-fitur saling independen serta memperhitungkan jumlah.

4.7 Evaluasi Analisis Sentimen

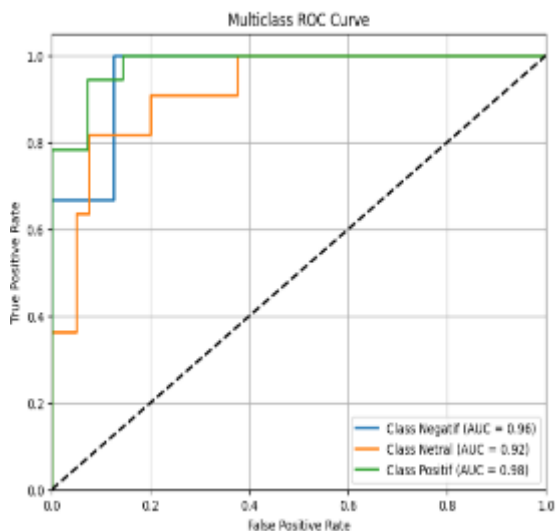
Pada tahapan ini akan dilakukan evaluasi hasil dari model pengujian digunakan yaitu *Confusion Matrix*, *Area Under Curve*, dan *Report* yang berisi akurasi, *Precision*, *Recall*, *F1-Score*.



Gambar 3. Hasil Pengujian Confusion Matrix

Confusion matrix digunakan sebagai cara menganalisis

sebuah kualitas data dari kelas positif, netral, dan negatif terhadap data aktual dan prediksi.



Gambar 4 Area Under Curve

4.8 Frekuensi Kata Dari hasil nilai AUC dari Naïve Bayes diperoleh AUC untuk kelas Positif sebesar 98%, kelas Netral sebesar 92%, dan kelas Negatif 96%. Nilai tersebut termasuk ke dalam *excellent classification*, dikarenakan nilai berada pada rentang 0.90 - 1.00.

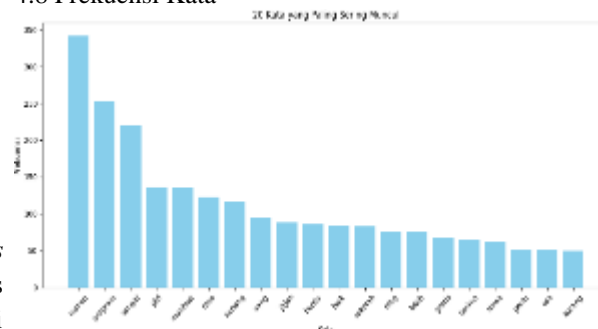
Tabel 4 Naïve Bayes Classification

Report	Positif	Netral	Negatif
Precision	0.92	0.75	0.50
Recall	0.97	0.55	0.67
F1-Score	0.95	0.63	0.57
Accuracy	86.27%		

Pada tabel *Naïve Bayes Classification* mendapatkan akurasi sebesar 86.27% secara keseluruhan model dapat mengklasifikasikan data dengan baik. Untuk kelas Positif *Precision* mendapatkan sebesar 92% ini menandakan bahwa model mampu memprediksi positif dengan benar, *Recall* sebesar 97% artinya model dapat menangkap data positif yang ada sebesar 97% sehingga sedikit kesalahan dalam mengenali kelas positif, *F1-Score* sebesar 95% kombinasi antara *precision* dan *recall* cukup tinggi, menunjukkan model tidak hanya akurat, tetapi juga konsisten dalam mendeteksi. Lalu kelas Netral dengan *Precision* mendapatkan sebesar 75% berarti bahwa dari prediksi yang termasuk kelas Netral

sesuai dengan label aslinya, lalu *recall* 55% menunjukkan model hanya berhasil mendeteksi setengah dari total data yang seharusnya diklasifikasikan sebagai netral, *F1-Score* sebesar 63% menunjukkan bahwa meskipun model cukup baik dalam menghindari *false positive*, masih banyak data netral yang tidak dikenali. Dan yang terakhir adalah kelas negatif dengan *precision* 50% ini hanya setengah dari prediksi negatif yang benar-benar sesuai, menandakan terdapat *false positive* yang cukup banyak, lalu nilai *recall* 67% menandakan model mampu mengenali sekitar dua pertiga dari data yang sebenarnya negatif, selanjutnya *F1-Score* 57% performa ini termasuk rendah cukup banyak data negatif yang tidak dikenali oleh model.

4.8 Frekuensi Kata



Gambar 5 Frekuensi Kata

Gambar diatas menunjukkan 20 kata yang paling sering muncul dalam respon siswa terhadap program makan bergizi gratis di SMA Negeri 01 Manokwari. Kata dengan angka tertinggi dengan adalah “makan”, yang muncul lebih dari 342 kali, disusul oleh kata “program” muncul sebanyak 252, “sangat” sebanyak 220, dan “gizi” sebanyak 146. data ini mengindikasikan bahwa siswa sangat setuju pada aspek inti dari program, yaitu kegiatan makan dan kualitas gizinya. Dalam respon dari siswa terdapat kata dukungan seperti kata-kata “uang”, “bantu”, “jajan” ini bukti bahwa siswa merasakan hal positif pada finansial mereka karena program MBG yang pemerintah berikan, namun terdapat kata-kata seperti “perlu”, “asa”, dan “kurang” ini juga bukti bahwa adanya masukan dan kritik dari siswa yang menerima program MBG yang harus diperbaiki ataupun ditambahkan. Keseluruhan hasil menunjukkan bahwa mayoritas dari siswa merespon dengan positif program makan gratis ini, namun tetap terdapat ruang untuk evaluasi dan peningkatan pada programnya.

5. Kesimpulan

5.1 Simpulan

Berdasarkan hasil penelitian analisis sentimen siswa SMAN 01 Manokwari terhadap efektivitas Program Makan Bergizi Gratis menggunakan algoritma *Naïve Bayes*, dapat disimpulkan bahwa mayoritas siswa memberikan tanggapan positif terhadap program tersebut, dengan persentase sentimen positif sebesar

74.2%, sentimen netral sebesar 15.1% dan sentimen negatif sebesar 10.7%. Proses klasifikasi menggunakan algoritma *Naïve Bayes* dengan metode *MultinomialNB* berhasil mencapai tingkat akurasi sebesar 86,27%. Hasil menunjukkan bahwa model bekerja cukup baik dalam mendeteksi sentimen positif, namun performanya masih perlu ditingkatkan dalam mengklasifikasikan sentimen netral dan negatif karena terdapat indikasi *overfitting*. Temuan ini menunjukkan bahwa mayoritas siswa memberikan tanggapan positif atau setuju terhadap program Makan Bergizi Gratis. Keterbatasan penelitian terletak pada jumlah data yang relatif terbatas, sehingga hasil yang diperoleh belum dapat digeneralisasi secara menyeluruh.

5.2 Saran

Untuk meningkatkan kualitas dan generalisasi hasil penelitian, disarankan agar jumlah data diperluas dengan melibatkan responden dari berbagai sekolah. Pelabelan data juga perlu ditingkatkan akurasinya melalui keterlibatan lebih dari satu annotator guna mengurangi subjektivitas, disertai pengukuran tingkat kesepakatan seperti *Cohen's Kappa*. Meskipun *Multinomial Naïve Bayes* menunjukkan performa baik pada sentimen positif, performa pada kelas netra dan negatif masih perlu ditingkatkan. Oleh karena itu, penelitian selanjutnya disarankan mengeksplorasi algoritma lain seperti *SVM*, *Random Forest*, atau model berbasis deep learning seperti *LSTM* dan *BERT*. Ketidakseimbangan data perlu diatasi dengan teknik seperti *SMOTE*, *oversampling*, atau *undersampling* agar model lebih mampu mengenali pola pada kelas minoritas. Selain itu, evaluasi model secara berkala dengan data terbaru penting untuk memantau perubahan sentimen siswa terhadap program.

Daftar Rujukan

- [1] M. Ali, R. Johnson, L. Smith, dan T. Nguyen, "Nutritional Interventions in Children and Their Impact on Academic Performance," *Journal of School Health*, vol. 92, no. 4, pp. 215–222, Apr. 2022, doi: 10.1111/josh.13121.
- [2] J. F. W. Cohen, A. A. Hecht, G. M. McLoughlin, L. Turner, dan M. B. Schwartz, "Impact of Universal Free School Meals on Student Dietary Quality and Academic Performance: A Systematic Review," *Public Health Nutrition*, vol. 24, no. 5, pp. 1152–1165, Mar. 2021, doi: 10.1017/S1368980020005044.
- [3] T. Cover dan P. Hart, "Nearest Neighbor Pattern Classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967, doi: 10.1109/TIT.1967.1053964
- [4] S. Hartati dan L. Kusumawardani, "Sentiment Analysis of COVID-19 Vaccine using Naïve Bayes Algorithm," *Procedia Computer Science*, vol. 179, pp. 705–712, 2021, doi: 10.1016/j.procs.2021.01.060.
- [5] E. M. A. Ernamia dan A. Herliana, "Analisis Sentimen Kuliah Daring dengan Algoritma *Naïve Bayes*, *K-NN* dan *Decision Tree*," *Jurnal Responsif: Riset Sains dan Informatika*, vol. 4, no. 1, hlm. 70–80, Feb. 2022.
- [6] J. C. Pamham, C. Millett, dan E. P. Vamos, "School meals in the UK: ultra-processed, unequal and inadequate," *Public Health Nutrition*, vol. 26, no. 1, pp. 297–301, Oct. 2022, doi: 10.1017/S1368980022002336.
- [7] J. R. Quinlan, "Improved Use of Continuous Attributes in C4.5," *Journal of Artificial Intelligence Research*, vol. 4, pp. 77–90, 1996, doi: 10.1613/jair.279.
- [8] J. D. Rennie, L. Shih, J. Teevan, dan D. R. Karger, "Tackling the poor assumptions of Naïve Bayes text classifiers," dalam *Proceedings of the 20th International Conference on Machine Learning (ICML 2003)*, Washington, DC, USA, 2003, pp. 616–623.
- [9] Z. Kastrati, F. Dalipi, A. S. Imran, K. Pireva Nuci, dan M. A. Wani, "Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study," *IEEE Access*, vol. 9, pp. 15242–15269, 2021, doi: 10.1109/ACCESS.2021.3056689.
- [10] B. Gunawan, H. S. Pratiwi, dan E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode *Naïve Bayes*," *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, vol. 4, no. 2, pp. 113–120, 2018, doi: 10.26418/jp.v4i2.27526.
- [11] A. Indriani, "Analisa Perbandingan Metode *Naïve Bayes Classifier* Dan *K-Nearest Neighbor* Terhadap Klasifikasi Data," *Sebatik*, vol. 24, no. 1, pp. 1–7, 2020, doi: 10.46984/sebatik.v24i1.909.
- [12] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, dan C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode *Linear Regression*," *Jurnal Nasional Teknologi Dan Sistem Informasi*, vol. 7, no. 1, pp. 8–17, 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.
- [13] B. Budiman, R. Nursyanti, R. Y. R. Alamsyah, dan I. Akbar, "Data Mining Implementation Using Naïve Bayes Algorithm and Decision Tree J48 In Determining Concentration Selection," *Jurnal XYZ*, vol. 1, no. 3, pp. [halaman], 2020.
- [14] Fitri, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma *Naïve Bayes*, *Random Forest* Dan *Support Vector Machine*," *Jurnal Transformatika*, vol. 18, no. 1, pp. 71–81, 2020, doi: 10.26623/transformatika.v18i1.23.
- [15] Z. S. Anggraeni, B. Budiman, C. Habibi, dan N. Alamsyah, "Analisis Sentimen Publik Media Sosial *Twitter* Terhadap *Tiket.com* Menggunakan Algoritma Klasifikasi," *Jurnal Informatika*, vol. 11, no. 1, pp. 3–4, 2024.
- [16] Agam, P. M., Ranius, M. I. A., Sutabri, T., & Ranius, A. Y. (2023). Sentiment Analysis pada Review Pengguna Aplikasi *Snapchat* dengan *Vader* dan Algoritma *Machine Learning*. *Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya*, 5(2), 94-97. DOI: 10.52303/jb.v5i2.1